

Hierarchical Visualization of Multidimensional Data

Lan Shu Xiangzhi He

School of Applied Mathematics, University of Electronic Science and Technology of China, Chengdu 610054, P. R. China

Abstract

Visualization of multidimensional data has been studied for a few decades and is still hot today. However, the proposed approaches have limitations when they meet multidimensional data. Hierarchical approach proposed in this paper is utilized to represent multidimensional data in multi-levels. First, the dimensions that have not been visualized are divided into two parts. Then the first part of the dimensions is visualized in a lower dimensional space. Subsequently, the second part of the dimensions is inspected in the next levels. Repeat the previous steps and thus the data are visualized hierarchically in a few recursions.

Keywords: Hierarchical visualization, Multi-dimensional data, Coordinate system

1. Introduction

One of the most direct approaches to understand data is to visualize them. Visualization of data reveals the basic but probably the most important information of the data and thus helps to make decisions according to the viewgraphs of the data. Visualization is limited to two dimensions of a screen or a piece of paper. The impact of visualization partly profits from the development of high-speed computers and high-resolution devices. On the other hand, the research of statistical data analysis, data mining and other related fields promotes the evolution of data visualization [1]-[3].

The visualization of discrete data is often plotted by symbols: dots, circles, stars, and so on. And the visualization of continuous data is often represented by curves or surfaces [4]. The visualization of continuous data is realized by visualizing discrete data in practice in a digital computer. The hierarchical approach proposed in this paper is suitable for visualizing both discrete and continuous data.

Data which are not more than three dimensions can be plotted on a 2-dimensional coordinate system easily. However, for higher dimensions, much work has to do to visualize data in 2-dimensional coordinate systems for some pre-presentation processes.

There are two basic ways to visualize multi-dimensional data on 2-dimensional coordinate systems. One is to use projections and plot data on multiple 2-dimensional views. Furnas and Buja discussed the usefulness of projections and sections together in identifying structures in data [5]. Tukey and Tukey put forward “draftsman’s display” and characterized the data in multiple views [6]. In the “grand tour”, a plane is moved through a d -dimensional scatter plot [7].

The other way is to use 2-dimensional graphical objects other than points to represent characters of the data. Star diagrams are one of the most typical approaches to visualize multidimensional data in this way. Chernoff faces use characterized human faces to represent multidimensional data [8].

This paper is organized as follows: Section 2 discusses the hierarchical visualization of multidimensional data. In Section 3, an example is represented to illustrate the approach discussed. The summary as well as a brief discussion and our future work are presented in Section 4.

2. Hierarchical visualization of multidimensional data

A d -dimensional datum $X = (X_1, X_2, \dots, X_d)$ is a d -tuple where X_i may be a discrete variable or continuous variable for $i = 1, 2, \dots, d$. More coordinates should be introduced to visualize X in a low dimensional coordinate system without losing any information. The hierarchical visualization of X discussed in this section divides the dimensions of X to be a few hierarchies and for each hierarchy the data of some dimensions are visualized with the left data visualized in the next hierarchies.

2.1. A Framework for hierarchical visualization of multi-dimensional data

Given a d -dimensional space Γ with the dimension set $D = \{\lambda_1, \lambda_2, \dots, \lambda_d\}$, let $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_k\}$ be a partition of D such that $|\mathcal{P}_i|$ -dimensional space can be represented in the coordinate system Γ_i . Then the

visualization of a point $X \in \Gamma$ can be visualized progressively. First, the data with respect to dimension set \mathcal{P}_1 are represented in the coordinate system Γ_1 . Then visualize the next $|\mathcal{P}_2|$ data. The recursion executes k times to visualize X without losing any information. The whole procedure can be illustrated by Fig. 1.

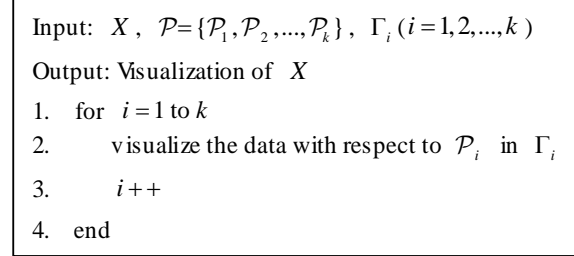


Fig. 1: Hierarchical visualization of data in d -dimensional spaces.

This algorithm treats the dimensions hierarchically. For each hierarchy, a few dimensions are considered while the dimensions that have not been considered are treated as a granule and they are left to be processed in the next hierarchies. And for each hierarchy, two layers can be considered to give insight into the hierarchy, that is, the outside layer and the inside layer. In the view of the outside layer, the dimensions that have not been displayed appear to be a point and in the view of the inside layer, the dimensions that have not been displayed appear to be a coordinate system.

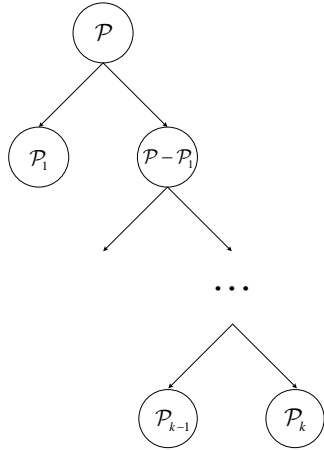


Fig. 2: Binary tree depiction of hierarchical visualization of data in d -dimensional space.

According to the algorithm illustrated by Fig. 1, a binary tree can be utilized to depict the visualization procedure and the binary tree corresponding to Fig. 1 is showed in Fig. 2.

The root of the binary tree denotes the point $X \in \Gamma$. All the left children as well as the last right child denote the visualization of the according dimensions and the right children but the last one denote that the visualization of the according dimensions will be showed in their successor nodes.

2.2. Two strategies of hierarchical visualization of multi-dimensional data

The key procedure of hierarchical visualization of multidimensional data is to choose the coordinate system for each \mathcal{P}_i . There are two strategies to determine the coordinate systems.

The first one is to choose the same kind of coordinate systems such as rectangular coordinate systems. Each coordinate system may be chosen to be

the same such that $|\mathcal{P}_i| = \left\lceil \frac{|\mathcal{P}|}{k} \right\rceil$ for $i = 1, 2, \dots, k$,

where $\left\lceil \frac{|\mathcal{P}|}{k} \right\rceil$ denotes the minimum integer which is not less than $\frac{|\mathcal{P}|}{k}$. For a d -dimensional datum X ,

$\left\lceil \frac{d}{2} \right\rceil$ coordinate systems should be utilized to visualize X if 2-dimensional coordinate systems are chosen to depict \mathcal{P}_i 's.

The second strategy does not constrain the coordinate systems that they should be the same kind. This strategy is more flexible as it allows kinds of coordinate systems to depict data together and the plots in a coordinate system may represent more dimensions. For example, the introduction of color brings in three more coordinate (RGB) to depict higher dimensional data [9].

3. An example for hierarchical visualization of multidimensional data

In this section, an example is stated to illustrate the approach of hierarchical visualization of multidimensional data. Some classic approaches of visualization multidimensional data such as star diagrams and Chernoff faces are represented to compare with that discussed in this paper.

The data in Table 1 is generated randomly. The strategies discussed above are adopted to visualize the data in Table 1. Fig. 3 and Fig. 4 are produced by the Matlab function *glyphplot* using the default settings

and they show the star diagrams and Chernoff faces of data in Table 1, separately.

Observation	x_1	x_2	x_3	x_4	x_5
1	3	5	6	5	2
2	8	0	6	1	0
3	7	7	6	6	0
4	8	8	3	0	7

Table 1: Data for the example.

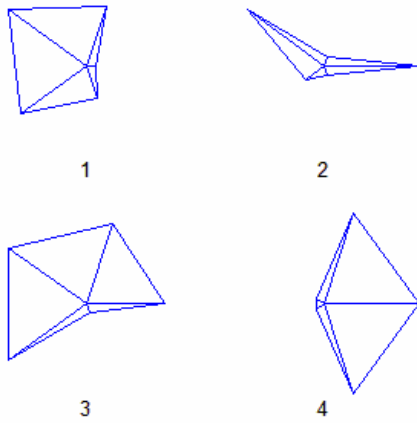


Fig. 3: Star diagrams of the data in Table 1.

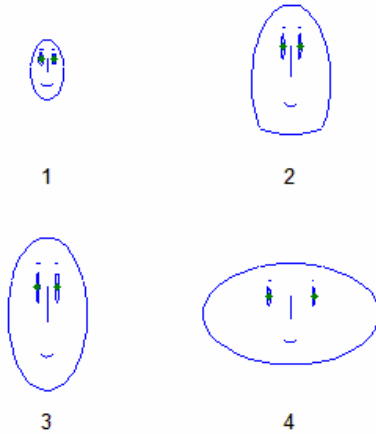


Fig. 4: Chernoff faces of the data in Table 1.

In Fig. 5, the first strategy is applied and two hierarchies with two coordinate systems work together to visualize the data in Table 1.

As it shows in Fig. 6, color is applied to denote the other 3 dimensions. The data has to be prepared in order to be represented by the color dimensions of RGB and each value is labeled in the parentheses nearby.

As shown in Fig. 3 to Fig. 6, star diagrams and Chernoff faces depict discrete multidimensional data

easily. However, they cannot depict continuous functions in multidimensional space. The visualization utilized the first strategy can depict continuous functions while too many coordinate systems may bring in some confusions. The second strategy performs best but may not visualize very high dimensional data as it is difficult to find suitable coordinate systems. In this situation, the two strategies should be integrated to show the data hierarchically.

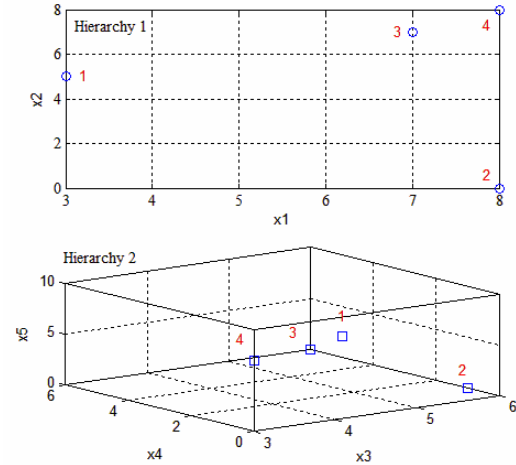


Fig. 5: Visualization of the data in Table 1 with the same kind of coordinate system.

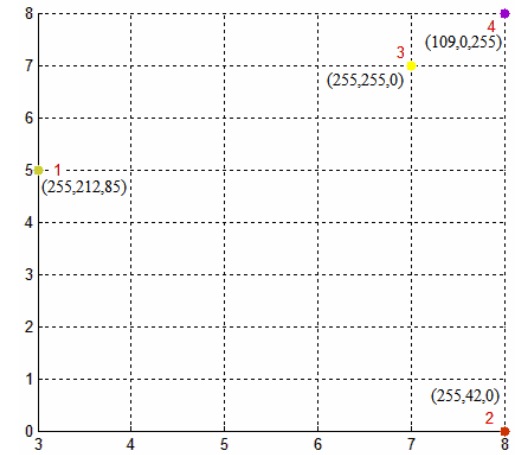


Fig. 6: Visualization of the data in Table 1 with color coordinate system.

4. Conclusions

The hierarchical approach discussed in this paper provides a new strategy to visualize multidimensional data. This approach also coincides with human's cognition of concepts, which is a kind of hierarchical structure, actually.

However, the hierarchical approach may take much time and space to visualize data in a 2-

dimensional space. Besides, finding the dimensions that similar to color dimensions may be difficult because not all dimensions of human's perception can be depicted in only one 2-dimensional space directly.

Our future work is to visualize complicated objects in a more convenient and acceptable way by the aid of computers. Moreover, the reduction of objects to be visualized is to be considered in order to get better visualization effects.

Acknowledgement

This work is partially supported by National Natural Science Foundation of China (Grant No. 10671030) and Fostering Plan for Young and Middle Age Leading Researchers in UESTC (Grant No. Y02018023601033).

References

- [1] G. Dzemyda, Multidimensional data visualization in the statistical analysis of curricula. *Computational Statistics & Data Analysis*, 49:265-281, 2005.
- [2] C.H. Chang and Z.-K. Ding, Categorical Data Visualization and Clustering Using Subjective Factors. *Data & Knowledge Engineering*, 53: 243-262, 2005.
- [3] D. Banu and A. Songul, Visualization and analysis of classifiers performance in multi-class medical data. *Expert Systems with Applications*, 34:628-634, 2008.
- [4] J.E. Gentle, *Elements of computational Statistics*, Springer-Verlag, New York, pp. 153-155, 2002.
- [5] G.W. Furnas and A. Buja, Prosection views: Dimensional inference through sections and projections. *Journal of Computational and Graphical Statistics*, 3:323-385, 1994.
- [6] P.A. Tukey and J.W. Tukey, Preparation; prechosen sequences of views. *Interpreting Multivariate Data*, pp. 189-213, 1981.
- [7] A. Daniel, The grand tour: a tool for viewing multidimensional data. *SIAM Journal on Scientific and Statistical Computing*, 6:128-143, 1985.
- [8] H. Chernoff, The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of American Statistic Association*, 68:361-368, 1973.
- [9] S.E. Fienberg, Graphical Methods in Statistics. *The American Statistician*, 33:165-178, 1979.